

Uma Breve Introdução à Estatística Bayesiana Aplicada ao Melhoramento Genético Animal



Empresa Brasileira de Pesquisa Agropecuária

Embrapa Amazônia Oriental

Ministério da Agricultura, Pecuária e Abastecimento

Documentos 412

Uma Breve Introdução à Estatística Bayesiana Aplicada ao Melhoramento Genético Animal

Vivian Dagnesi Timpani

Thialla Emille Costa do Nascimento

Embrapa Amazônia Oriental

Tv. Dr. Enéas Pinheiro, s/n. CEP 66095-903 – Belém, PA.

Caixa Postal 48. CEP 66017-970 – Belém, PA.

Fone: (91) 3204-1000

Fax: (91) 3276-9845

www.embrapa.br

www.embrapa.br/fale-conosco/sac

Comitê Local de Publicação

Presidente: *Silvio Brienza Júnior*

Secretário-Executivo: *Moacyr Bernardino Dias-Filho*

Membros: *Orlando dos Santos Watrin*

Eniel David Cruz

Sheila de Souza Correa de Melo

Regina Alves Rodrigues

Luciane Chedid Melo Borges

Supervisão editorial: *Luciane Chedid Melo Borges*

Revisão de texto: *Narjara de Fátima Galiza da Silva Pastana*

Normalização bibliográfica: *Andréa Liliane Pereira da Silva*

Tratamento de imagens, ilustração da capa e editoração eletrônica: *Vitor Trindade Lôbo*

1ª edição

On-line (2015)

Disponível em: www.embrapa.br/amazonia-oriental/publicacoes

Todos os direitos reservados

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Amazônia Oriental

Timpani, Vivian Dagnesi.

Uma breve introdução à estatística bayesiana aplicada ao melhoramento genético animal / Vivian Dagnesi Timpani, Thialla Emille Costa do Nascimento. – Belém, PA : Embrapa Amazônia Oriental, 2015.

57 p. : il. ; 15 cm x 21 cm. – (Documentos / Embrapa Amazônia Oriental, ISSN 1983-0513; 412).

1. Estatística. 2. Melhoramento genético animal. 3. Genética animal. I. Nascimento, Thialla Emille Costa do. II. Título. III. Série.

CDD 21. ed. 519.542

© Embrapa 2015

Autores

Vivian Dagnesi Timpani

Zootecnista, doutora em Zootecnia, pesquisadora da Embrapa Amazônia Oriental, Belém, PA.

Thialla Emille Costa do Nascimento

Médica-veterinária, graduada pela Universidade Federal do Vale do São Francisco, Petrolina, PE.

Agradecimentos

Ao Professor Dr. Tarcisio Moraes Gonçalves (Ufla).

À Professora Dra. Thelma Safadi (Ufla).

Ao Professor Dr. Idalmo Garcia Pereira (UFMG).

Ao Professor Blasco.

Apresentação

Depois da análise dos dados coletados no experimento ou campo, o propósito de qualquer pesquisador é fazer inferências ou previsões, com certo grau de confiança, sobre o fenômeno estudado. A ciência da estatística apresenta vários métodos com essa finalidade e a estatística bayesiana vem nos últimos anos ganhando força em algumas áreas, como no melhoramento genético animal, sendo uma metodologia alternativa para a avaliação genética de rebanhos comerciais. Este material é um apanhado de vários trabalhos da literatura e foi feito com o intuito de resumir a teoria Bayesiana, de maneira simples e prática, e apresentar sua aplicação no melhoramento genético de animais e plantas.

Adriano Venturieri

Chefe-Geral da Embrapa Amazônia Oriental

Sumário

Uma Breve Introdução à Estatística Bayesiana Aplicada ao Melhoramento Genético Animal	11
Introdução.....	11
Estatística clássica x Estatística bayesiana.....	13
Probabilidade: o paradigma bayesiano.....	18
Fundamentos da inferência bayesiana.....	22
Teorema de Bayes.....	22
Informação <i>a priori</i>	27
Escolha da distribuição <i>a priori</i>	28
Tipos de distribuição <i>a priori</i>	31
<i>Priori</i> própria.....	31
<i>Priori</i> não informativa ou vaga	32
Distribuições <i>a priori</i> usadas no melhoramento animal.....	33
Inferência bayesiana.....	34

Variável Aleatória.....	34
Densidade de Probabilidade	36
Intervalo de credibilidade	37
Distribuição <i>a posteriori</i>	38
Distribuições que compõem a distribuição <i>a posteriori</i>	39
Distribuição de probabilidade conjunta	39
Distribuição de probabilidade marginal	39
Distribuição de probabilidade condicional	40
Obtenção do <i>posteriori</i> na estatística bayesiana	40
Método de Monte Carlo via cadeia de Markov (MCMC)	40
Algoritmo Metropolis-Hastings	44
Amostrador de Gibbs.....	45
Diagnóstico de convergência	47
Critérios de seleção do modelo e hipóteses para análise estatística.....	48
Fator de Bayes	49
Princípios da estatística bayesiana no modelo animal	52
Modelo animal	52
Estimação dos componentes de variância	53
Estimação dos valores genéticos e seleção.....	54
Referências	55

Uma Breve Introdução à Estatística Bayesiana Aplicada ao Melhoramento Genético Animal

Vivian Dagnesi Timpani

Thialla Emille Costa do Nascimento

Introdução

Quando comecei a estudar a estatística bayesiana em virtude da minha tese de doutorado, como zootecnista, senti uma falta muito grande de literatura que explicasse não o “matematiquês” do assunto, mas sim os princípios, definições e interpretações que deveriam ser entendidos e usados. Quando me propuseram a realizar um curso sobre o assunto, pensei: por que não transformar toda aquela estatística dos livros de estatística bayesiana em um material que pudesse ser utilizado por profissionais da área de Agrárias com simplicidade? Assim, coloquei-me na frente do meu computador e de uma montanha de livros e artigos para produzir este material, que é um apanhado mais um resumo de vários trabalhos, em que tentei traduzir a beleza e funcionalidade dessa teoria.

Foi então que tive a oportunidade de ler um livro muito interessante, chamado *Uma senhora toma chá...*, que iniciava contando uma história. Diz a história que, numa tarde de verão na Inglaterra, um grupo de professores tomava chá (lembrem-se: os ingleses tomam chá com leite todas as tardes) e uma senhora afirmou que o chá servido sobre o leite tinha um gosto diferente do leite servido sobre o chá. Um homenzinho (porque ele era baixinho!), chamado Ronald Fischer (com certeza

você já ouviram falar nele), interessou-se pelo problema e tentou entendê-lo, publicando, em 1953, o livro *The design of experiments*, que ajudou a revolucionar a experimentação em todos os campos da Ciência na primeira metade do século 20. Nesse livro, Fischer analisa a crença no problema proposto pela senhora do chá e considera as várias possibilidades de planejar o experimento, desenvolvendo modelos estatísticos para obter um resultado que pudesse ser interpretado, confirmando ou não, a diferença no gosto da bebida (SALSBURG, 2009).

A importância de Fischer para a experimentação agropecuária está no fato de ele ter sido o primeiro a mostrar (na forma de publicação) a aplicação da estatística, de forma simples e resumida, por meio da análise de variância, influenciando os trabalhos científicos realizados no século 20 (SALSBURG, 2009). Esse modo de pensar na estatística fez com que Fischer – considerado o criador dos métodos modernos de análise e experimentação – desenvolvesse juntamente com Karl Pearson – outro grande nome da estatística, que estudou as distribuições de frequência – a teoria da correlação, o teste X^2 , dentre outras contribuições (MEMÓRIA, 2004).

Fischer também trabalhou com o conceito de probabilidade e, nesse assunto, encontrou alguns adversários que desenvolveram o conceito de probabilidade inversa (ou, como chamamos na estatística bayesiana, probabilidade *a posteriori*) com base nos estudos do Reverendo Thomas Bayes.

A metodologia bayesiana surgiu em consequência da publicação *An essay towards solving a problem in the doctrine of chances*, atribuída ao Reverendo Thomas Bayes e comunicado à Royal Statistical Society após sua morte por Richard Price, em 1763 (BAYES, 1763), e posteriormente reproduzido na revista *Biometrika*.

Segundo Blasco (2008), o uso da probabilidade *a posteriori* como método de estimação é mais eficiente que outros métodos de inferência estatística, o que torna a estatística bayesiana bastante interessante.

A estatística bayesiana foi aplicada no melhoramento genético animal pela primeira vez em 1986, por meio de um artigo publicado no *Journal of animal science* por Daniel Gianola e Rohan Fernando, no qual propuseram a utilização da estatística bayesiana no melhoramento genético animal pelo fato de que as pressuposições do método clássico (normalidade e não viesado) poderiam ser irrelevantes para a classificação dos animais submetidos à seleção.

Estatística clássica x Estatística bayesiana

Depois da análise dos dados, o propósito de qualquer estatístico é fazer inferências ou predições, com certo grau de confiança, sobre o fenômeno que se está estudando, a partir dos dados que representam a variabilidade ou a incerteza na observação da característica ou fenômeno.

Inferência significa operação intelectual por meio da qual se afirma a verdade de uma proposição em decorrência de sua ligação com outras já reconhecidas como verdadeiras (DICIONÁRIO..., 2009).

Contudo, na “Ciência Estatística” podemos encontrar diversas escolas, ou seja, “diferentes maneiras de pensar” sobre como analisar os fenômenos, com princípios e procedimentos próprios de cada metodologia (PAULINO et al., 2003).

Os procedimentos da estatística bayesiana podem ser melhor compreendidos quando comparados à sua principal “concorrente”, a estatística clássica.

Na estatística clássica, as inferências sobre um fenômeno ou característica que ocorre na população estudada é baseada na avaliação dos parâmetros estimados de amostras retiradas dessa população. Assim, ocorre uma generalização dos parâmetros da população, a partir da estimação dos parâmetros das amostras obtidas dessa população. Nesse caso, a interpretação dos dados não depende apenas das amostras, mas também das hipóteses adotadas sobre o fenômeno estudado (PAULINO et al., 2003).

Amostras são observações, dentre as várias possíveis de serem obtidas na população, por meio da repetição do experimento.

População é o conjunto de todas as observações possíveis do fenômeno ou característica estudados.

Para Pearson, todos os valores que medimos de um determinado fenômeno são, na realidade, parte de uma dispersão aleatória descrita por uma probabilidade de acontecerem, que ele chamou de função de distribuição.

Parâmetros são medidas usadas para descrever o fenômeno ou a característica estudada.

Distribuição Normal ocorre quando a média é a probabilidade de o fenômeno ocorrer, ou seja, são os valores de maior frequência. Neste caso, a média é o centro da distribuição.

Segundo Salsburg (2009), cada função de distribuição (que na maioria dos métodos da estatística clássica pressupõe-se que seja Distribuição Normal) pode ser descrita por quatro parâmetros:

- **Média:** valor ou posição central a partir do qual as medições se dispersam.
- **Desvio-padrão:** o quanto a maioria das medições se dispersam em torno da média.
- **Simetria:** grau em que as medições se acumulam em apenas um dos lados da média.
- **Curtose:** quanto as medidas extremas se afastam da média.

OBSERVAÇÃO

Nunca podemos determinar realmente esses parâmetros; podemos apenas estimá-los pelos dados.

As inferências baseiam-se no número de estimativas que estão distribuídas ao redor do valor parâmetro e espera-se que o intervalo contenha, em certa proporção das amostras (repetições), o valor real do parâmetro. Dessa forma, o resultado é dado em probabilidade e a estimativa do parâmetro em estudo é uma distribuição de densidade probabilística, em um determinado conjunto de dados (BLASCO, 2001).

Na metodologia bayesiana, o conceito de repetição não existe. A informação do parâmetro estudado é resumida pela maior probabilidade de o parâmetro assumir um determinado valor ou pela probabilidade de o valor do parâmetro estar dentro de certo intervalo (BLASCO, 2001; FARIA et al., 2007).

Para a estatística clássica, o valor real do parâmetro é sempre fixo e as repetições ou amostras são aleatórias e a incerteza sobre o valor real do parâmetro (erro) é ignorada (BLASCO, 2001).

Para a estatística bayesiana, a amostra (conjunto de dados) é fixa e o parâmetro é aleatório (BLASCO, 2001; FARIA et al., 2007). A inferência bayesiana tem como característica essencial o uso da probabilidade para quantificar as incertezas, ou seja, o grau de confiança que o pesquisador tem sobre o valor do parâmetro (SAVIAN, 2008).

A estatística clássica é representada, atualmente, no processo de seleção das espécies animais de interesse zootécnico, pelo modelo misto. No entanto, essa metodologia tem apresentado algumas limitações, tais como a necessidade de grande recurso computacional, a pressuposição da normalidade (nem todas as características avaliadas têm esse tipo de distribuição) e a construção de intervalos de confiança

corretos (NOGUEIRA et al., 2003). Outra limitação importante está no fato de que se obtêm apenas alguns parâmetros da distribuição dos dados, tais como média, mediana e moda, que podem ser irrelevantes para a estimação dos componentes de variância e dos parâmetros genéticos (GIANOLA; FERNANDO, 1986).

Algumas dessas limitações podem ser contornadas pela estatística bayesiana:

- Por tratar os parâmetros como variável aleatória, ela permite obter intervalos de confiança utilizando-se métodos iterativos (método de Monte Carlo), quando a distribuição das amostras impedir a construção dos intervalos de confiança (NOGUEIRA et al., 2003).
- No contexto bayesiano, o conceito do “viés”, que produz estimativas viciadas na estatística clássica, não existe porque o conceito de repetição do experimento não é aplicável a essa metodologia (BLASCO, 2001).

Viés: erro que ocorre tendenciosamente no procedimento estatístico, conduzindo a uma conclusão errônea sobre o fenômeno estudado.

- A metodologia bayesiana requer um número menor de observações, pois o uso de probabilidade na inferência diminui a dependência das estimativas dos parâmetros em relação ao número de observações utilizadas (SILVA, 2006).
- A inferência bayesiana tem a vantagem adicional de permitir a incorporação de informações anteriores (*a priori*) adquiridas antes da obtenção dos dados existentes sobre o parâmetro estudado, enriquecendo o processo de inferência (NOGUEIRA et al., 2003).

QUADRO COMPARATIVO

Estatística clássica x Estatística bayesiana

1. A estatística bayesiana utiliza o conhecimento prévio do pesquisador (informação *a priori*) adicionalmente aos dados na análise estatística. A informação *a priori* na estatística clássica não é usada porque é considerada, na maioria dos casos, subjetiva. A estatística clássica tem menos informações disponíveis para a análise, já a estatística bayesiana tem conclusões mais fortes para o mesmo conjunto de dados analisado.
2. Na estatística clássica, é considerado que há apenas um valor para o parâmetro estudado. Na estatística bayesiana, considera-se a existência de vários valores possíveis para o parâmetro, originando assim uma distribuição de probabilidades desses valores para os dados estudados.
3. Na estatística clássica, as inferências são feitas com base na verossimilhança e tratam os parâmetros como fixos e desconhecidos e os dados como aleatórios e conhecidos. Na estatística bayesiana, as inferências são provenientes da distribuição *a posteriori* (conhecimento prévio do pesquisador + dados do experimento) e os parâmetros são aleatórios e desconhecidos; e os dados são fixos e conhecidos.
4. Ambas as escolas, clássica e bayesiana, utilizam a verossimilhança na realização de suas inferências. De um lado, a verossimilhança assume um papel central na inferência bayesiana, como a função que expressa toda a informação proveniente dos dados. Por outro lado, o método da Máxima Verossimilhança ou ML (*Maximum Likelihood*) possui interessantes propriedades usadas pelos estatísticos clássicos.
5. No melhoramento genético animal, as inferências obtidas em ambas as escolas não são sempre coincidentes, particularmente para amostras pequenas e quando a análise bayesiana utiliza informações *a priori* muito informativas. Porém, quando o conjunto de dados é suficientemente grande, os resultados são bastante similares na maioria dos casos.

Probabilidade: o paradigma bayesiano

Quando um pesquisador planeja um experimento, ele deve levar em consideração as informações que existem nos dados utilizados e a presença de um elemento de incerteza ou erro (FERREIRA, 2005).

Paradigma: Termo com o qual Thomas Kuhn designou as realizações científicas que geram modelos que, por período mais ou menos longo e de modo mais ou menos explícito, orientam o desenvolvimento posterior das pesquisas exclusivamente na busca da solução para os problemas por elas suscitados (DICIONÁRIO...,2009).

O conceito de probabilidade na estatística bayesiana é extremamente importante e está ligado à ideia de grau de confiança (afinal o erro está interferindo na análise) ou veracidade de uma hipótese testada (PAULINO et al., 2003).

Hipótese: Suposição que orienta uma investigação por antecipar características prováveis do fenômeno estudado.

Sob o ponto de vista da estatística clássica, a definição mais interessante sobre probabilidade diz que a “Probabilidade de um dado evento é a proporção de vezes que esse evento ocorre em um determinado período de tempo” (SALSBURG, 2009). Em outras palavras, um evento (por exemplo, tirar 6 no lançamento de um dado) tem uma probabilidade x ligada ao número de vezes que o evento (número 6) aparece numa sequência de lançamentos seguidos do dado.

O conceito de probabilidade de um fenômeno é obtido realizando um experimento n vezes e anotando o número de vezes (r) que o fenômeno de interesse ocorre (FERREIRA, 2005).

$$P(A) = r/n$$

Onde:

r é o resultado obtido da ocorrência do evento estudado.

n é o número de repetições realizadas no experimento.

Do ponto de vista bayesiano, todos os parâmetros estudados em um fenômeno são desconhecidos e incertos e essa incerteza deve ser quantificada em termos de probabilidade (PAULINO et al., 2003; SORENSEN; GIANOLA, 2002).

Assim, a probabilidade representa o grau de crença (no caso, a opinião do pesquisador) condicionado à informação existente sobre o fenômeno estudado. É uma medida subjetiva e pode variar de pesquisador para pesquisador, pois a experiência e a fonte dessa informação que cada um possui são diferenciadas.

OBSERVAÇÃO

Essa informação prévia que o pesquisador possui sobre o fenômeno na estatística bayesiana é chamada de informação *a priori*.

Para os estatísticos bayesianos, cada problema estudado é único, com características e valores para os parâmetros estimados exclusivos. Com isso, o grau de conhecimento varia de problema para problema (mesmo que o objetivo de cada problema seja o mesmo, como por exemplo, estimar herdabilidade para peso a desmame na raça Nelore).

Assim, a informação *a priori* possui uma variabilidade que só pode ser interpretada em termos de distribuição de probabilidade, chamada de distribuição *a priori* (PAULINO et al., 2003).

No entanto, essa informação *a priori* está condicionada a certas condições experimentais de onde ela foi obtida. Segundo Salsburg (2009), a informação *a priori* pode ser explicada por meio da seguinte analogia:

...No século 18, a cidade de Veneza era governada por um grupo de famílias nobres que tinha como chefe de estado um membro de uma destas famílias – “doge” – que era escolhido por eleição. Nesta eleição, inicialmente, um pequeno grupo de membros das famílias reinantes era escolhido como eleitor. Esta escolha era feita através de um sorteio, onde, apenas, nove membros (dos 30 que compunham o grupo) seriam aceitos como eleitor e assim poderiam votar. Eram usadas 30 bolas de cera, com e sem uma tira de papel escrito eleitor dentro. Cada membro do grupo de candidatos a eleitor tirava (um por vez) uma bola.

Antes que o primeiro candidato tirasse a sua bola, a chance (probabilidade) de cada membro do grupo ser eleitor era de $9/30$. Se a primeira bola estivesse com o papel em branco, cada um dos remanescentes tinha a probabilidade $9/29$ de ser eleitor. Porém, se a primeira bola estivesse com a tira de papel escrita, cada um dos remanescentes tinha a probabilidade $8/29$ de ser eleitor. A cada bola tirada, a chance de o próximo membro ser eleito, diminuía ou aumentava, dependendo do resultado obtido anteriormente.

Esse é um exemplo de probabilidade condicionada aos eventos, ou como chamamos na estatística, probabilidade condicional. Assim, a probabilidade de cada membro se tornar eleitor dependia das bolas que tivessem sido sorteadas antes.

A probabilidade condicional é a probabilidade de ocorrer o evento B (depois), tendo ocorrido o evento A (antes).

O aspecto importante da probabilidade condicional é que a probabilidade do evento A é diferente para condições anteriores diferentes, ou seja, a probabilidade do evento A está condicionada aos eventos que ocorreram antes do evento que está sendo examinado.

Esta era a ideia que reinava entre os estatísticos clássicos do século 18 sobre como trabalhar com a probabilidade (SALSBURG, 2009).

OBSERVAÇÃO

A ideia da probabilidade condicional está baseada no fato de que um evento pode assumir diferentes valores e todos eles estarem corretos (FERREIRA, 2005).

Assim, a probabilidade condicional representa, em termos probabilísticos, a possibilidade de os eventos serem observados quando se impõem condições para que o experimento ou as análises sejam realizadas (SORENSEN; GIANOLA, 2002).

Porém, na última década do século 18, o Reverendo Thomas Bayes descobriu que poderíamos calcular a probabilidade do evento B (antes) condicionada à probabilidade do evento A (depois). Bayes inverteu o conceito de probabilidade e essa teoria foi chamada de “probabilidade inversa”.

Um exemplo para ajudar a entender a ideia de Bayes:

Probabilidade clássica: Qual a probabilidade de um fumante ter câncer no pulmão?

— DO ANTES PARA O DEPOIS —
DA CAUSA PARA O EFEITO

Probabilidade de Bayes: Qual a probabilidade de um paciente com câncer no pulmão ser fumante?

— DO DEPOIS PARA O ANTES —
DO EFEITO PARA A CAUSA

Fundamentos da inferência bayesiana

Teorema de Bayes

Com base na probabilidade condicional, a inferência bayesiana utiliza o Teorema de Bayes como uma ferramenta de atualização da opinião do pesquisador (SILVA, 2006) por meio da aprendizagem com a experiência. Em outras palavras, o conhecimento inicial sobre o parâmetro estudado é modificado depois que o experimento foi realizado (PAULINO et al., 2003; SAVIAN, 2008).

Em termos de probabilidade, esse teorema mostra como essas probabilidades mudam quando novos conhecimentos são usados na análise, pois essa informação está disponível antes e separadamente dos dados obtidos pelo experimento (LEANDRO, 2001).

$$P(B_i|A) = \frac{P(B_i|A) \times P(A|B_i)}{\sum P(B_j) \times P(A|B_j)}$$

onde:

B_i é o evento estudado no experimento.

A é o conjunto de repetições realizadas no experimento.

B_j são todos os eventos possíveis de ocorrer no experimento.

Porém, esse teorema é mais comumente representado pela expressão:

$$p(\theta|x) \propto p(x|\theta) p(\theta)$$

onde:

$p(\theta|x)$ é a distribuição a posteriori.

$p(\theta)$ é a distribuição a priori do parâmetro.

$p(x|\theta)$ é a função de verossimilhança obtida pelos dados (condicional).

Ou seja, o Teorema de Bayes pode ser interpretado como:

“A probabilidade a posteriori é condicionada à probabilidade a priori multiplicada pela verossimilhança.”

Nesse teorema, há duas fontes de informações disponíveis, que são combinadas para obter como resultado a distribuição *a posteriori* $[p(\theta|x)]$ de onde serão realizadas as inferências do parâmetro estudado. A primeira é a distribuição *a priori* $[p(\theta)]$ do parâmetro, que expressa o conhecimento prévio do pesquisador, em termos probabilísticos. A segunda é a função de verossimilhança $[p(x|\theta)]$, que se refere aos dados amostrais obtidos no experimento (PAULINO et al., 2003; SAVIAN, 2008; SILVA, 2006).

Como x representa a distribuição dos dados, ele não é uma função ligada ao parâmetro θ , então a probabilidade *a posteriori* $[p(\theta|x)]$ é proporcional ao produto distribuição *a priori* $[p(\theta)]$ multiplicado pela função de verossimilhança $[p(x|\theta)]$ (POLICARPO, 2001).

No Teorema de Bayes, a probabilidade *a priori* é baseada nas informações disponíveis separadamente do experimento. Representa o conhecimento do fenômeno antes de os dados serem observados. A verossimilhança é deduzida dos dados e expressa todo o conhecimento do fenômeno contido nestes dados (LEANDRO, 2001).

Exemplo:

Existe um tipo de rato de laboratório, cuja cor da pele é controlada por um único gene com dois alelos 'A' e 'a', de modo que, quando o rato tem duas cópias do alelo recessivo (aa) a sua pelagem é marrom e, nos outros casos, a pelagem é preta (AA e Aa).

Se considerarmos uma situação em que um rato preto, filho de dois ratos pretos heterozigotos (Aa), é acasalado com um rato marrom (aa) e que deste cruzamento nasçam sete ratos pretos (A₋), pergunta-se:

Qual a probabilidade *a posteriori* de esse rato preto ser homozigoto (AA)?

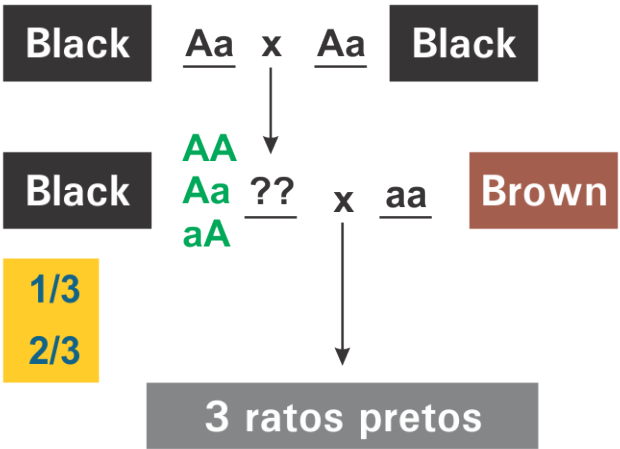


Figura 1. Esquema de cruzamento utilizado no exemplo.
Fonte: Blasco (2008).

Neste caso, após todos os cruzamentos possíveis entre o rato preto avaliado e outro marrom, a descendência e as probabilidades de cada genótipo ocorrer são apresentadas na Tabela 1.

Tabela 1. Probabilidades *a priori* da descendência obtidas nos possíveis cruzamentos entre um rato preto e outro marrom.

Cruzamento	Preto (AA)	Preto (Aa)	Marrom (aa)
AA x aa	0	1	0
Aa x aa	0	1/2	1/2
Aa x Aa	1/4	1/2	1/4

Fonte: Leandro (2001).

Qual o impacto da informação *a priori* sobre o resultado deste experimento?

Isso é importante, pois as circunstâncias em que se realizam os experimentos são diferentes, mesmo que o objetivo seja o mesmo. Dificilmente um pesquisador consegue repetir, no mesmo grau e intensidade, todos os efeitos que influenciam o evento estudado, principalmente no caso do melhoramento animal.

Os efeitos incluídos no modelo mudam, as condições de ambiente mudam, a balança de pesagem muda, a pessoa que coleta os dados muda, o animal fica doente, falta ração em um determinado dia e tudo isso influencia de maneira significativa na coleta das observações.

Assim, as *prioris* devem ser avaliadas sempre, a cada experimento realizado.

ENTENDENDO O TEOREMA DE BAYES

$\theta \rightarrow$ parâmetro estudado: genótipo do rato preto.

θ $\left\{ \begin{array}{l} \text{é } 0 \text{ se o genótipo do rato preto for } AA. \\ \text{é } 1 \text{ se o genótipo do rato preto for } Aa. \end{array} \right.$

• Probabilidade *a priori* de θ :

Se $\theta = 0 \rightarrow P(AA) = 1/3$

Se $\theta = 1 \rightarrow P(Aa) = 2/3$

Existem duas situações para essa progênie, dependendo do genótipo do rato preto testado, e devemos calcular a probabilidade conjunta (Ver tabela 1 – valores grifados e em negrito):

$$P(y=3 \text{ black}) = P(y=3 \text{ black} \mid AA) \cdot P(AA) + P(y=3 \text{ black} \mid Aa) \cdot P(Aa) = 1 \cdot 1/3 + (1/2)^3 \cdot 2/3 = 0.42$$

• Pelo Teorema de Bayes:

Juntando a informação da *priori* com a informação obtida no cruzamento-teste, teremos:

$$P(AA \mid y=3 \text{ black}) = \frac{P(y=3 \text{ black} \mid AA) \cdot P(AA)}{P(y=3 \text{ black})} = \frac{1 \times 0.33}{0.42} = 0.80$$

$$P(y=3 \text{ pretos}) = 1$$

$$P(AA) = 1/3 = 0,33$$

$$P(Aa \mid y=3 \text{ black}) = 1 - P(AA \mid y=3 \text{ black}) = 1 - 0.80 = 0.20$$

$$P(y=3 \text{ pretos}) = 1$$

$$P(Aa) = 2/3 = 0,77$$

$$\text{prior } P(AA) = 0.33$$

$$\text{prior } P(AA \mid y) = 0.80$$

$$\text{prior } P(Aa) = 0.67$$

$$\text{prior } P(Aa \mid y) = 0.20$$

Com esse exemplo, observa-se que houve uma modificação nas probabilidades iniciais estabelecidas pela teoria genética, ou seja, houve uma atualização dos valores do parâmetro estudado quando se adicionou os dados em conjunto com o conhecimento prévio.

Observa-se também que a função de verossimilhança desempenha um papel importante no teorema: é por meio dela que o conhecimento é modificado. Ela é a representação das informações que os dados têm a contar sobre o parâmetro estudado (LEANDRO, 2001) e a distribuição *a posteriori* são os valores dos parâmetros atualizados após a realização do experimento.

Informação *a priori*

A informação *a priori* (expressa em termos de distribuição de probabilidade) é a informação sobre os parâmetros estudados obtidos ou existentes antes da realização do experimento e serve, na estatística bayesiana, para expressar tanto conhecimento como ignorância sobre esses parâmetros (LEANDRO, 2001).

Quando um fenômeno é estudado, várias hipóteses sobre os valores dos parâmetros podem ser formuladas pelo pesquisador. A informação *a priori* é a hipótese que o pesquisador fixa como sendo o valor verdadeiro do parâmetro estudado. Essa *priori* pode ser extraída de fundamentos subjetivos, considerações particulares ou informações anteriores disponíveis na literatura. Em outras palavras, a informação *a priori* pode ser interpretada como o grau de convicção sobre o valor do parâmetro estudado, antes do experimento ou observação dos dados, e esse conhecimento pode ser formalmente incorporado na análise (SORENSEN; GIANOLA, 2002).

O uso desse tipo de informação traz à tona uma questão que todo pesquisador deve fazer antes de iniciar suas análises.

Escolha da distribuição *a priori*

Sob o paradigma Bayesiano, o resultado de qualquer análise estatística combina a informação fornecida pelos dados com a informação *a priori* (DIAS; TOMAZELLI, 2011). Dependendo do problema estudado, há uma quantidade significativa de evidências “objetivas” na qual se pode basear a escolha da distribuição *a priori* utilizada.

Com muita informação disponível (vinda de fontes objetivas e confiáveis), o pesquisador pode assumir certa distribuição *a priori* consistente, que poderá contribuir com o resultado (afinal é para isso que a distribuição *a priori* serve) e essa distribuição deve representar, probabilisticamente, o conhecimento que se tem sobre os parâmetros antes da realização do experimento.

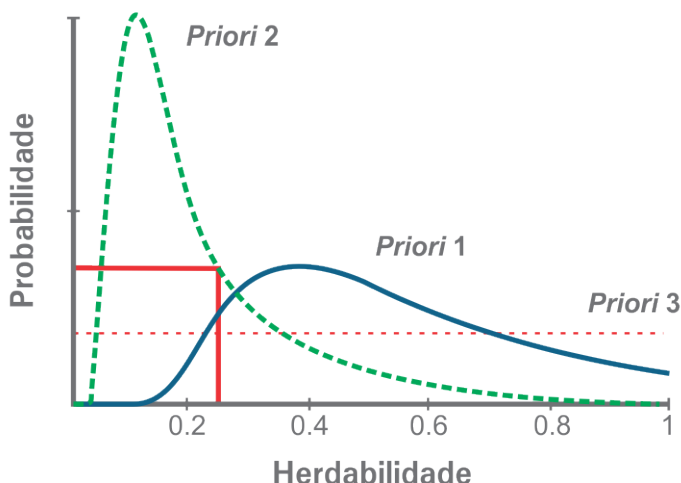


Figura 2. Diferentes *prioris* para herdabilidade da taxa de ovulação em suínos da raça Landrace, mostrando as curvas de densidade da probabilidade para cada uma delas.

Fonte: Blasco (2008).

As curvas 1 e 2 são assimétricas, pois a herdabilidade para essa característica é baixa. A curva 3 é uma *priori* que não possui uma informação relevante sobre a herdabilidade de taxa de ovulação em suínos.

Quando a informação da distribuição *a priori* é precisa, não há problemas para utilização da estatística bayesiana. Porém, a escolha da *priori* fica mais difícil e interessante em problemas nos quais não existe uma base teórica “sólida”, ou seja, ligada ao conhecimento oriundo de fontes confiáveis. Nessa situação, a *priori* utilizada é resultado único baseado no conhecimento “subjetivo” do pesquisador, que transmite seu “sentimento” ou opinião em relação ao valor do parâmetro estudado (BLASCO, 2008; LEANDRO, 2001).

Nesse caso, o princípio da coerência tem que prevalecer.

No entanto, mesmo que o pesquisador esteja satisfeito com a sua escolha, pode ser que a sua expectativa em relação à distribuição *a posteriori* seja frustrada em virtude do resultado encontrado para o parâmetro.

Portanto, escolher a *priori* é um passo importante na análise bayesiana.

Outro ponto que deve ser avaliado é a quantidade de informação que os dados utilizados podem trazer. A contribuição da *priori* é relativamente mais importante quando a quantidade de dados é escassa em comparação com quando é abundante, pois, em arquivos de dados grandes, a informação da verossimilhança domina o resultado obtido em relação à *priori* (ASSIS, 2002).

A questão primordial na escolha da *priori* está ligada ao fato de como obter e quantificar essa informação, de natureza essencialmente subjetiva, de modo a transformá-la em uma informação que possa ser utilizada (PAULINO et al., 2003).

OBSERVAÇÃO

Como a distribuição *a priori* influencia no resultado ou distribuição *a posteriori*: ver exemplo do fumante no item “Probabilidade: o paradigma bayesiano”.

Existem alguns métodos para se fazer essa transformação:

a) Método estrutural

Nesse método, a escolha da *priori* é baseada em questões relacionadas diretamente ao parâmetro estudado. Em outras palavras, é uma pesquisa que se realiza sobre o parâmetro estudado. Essas questões serão usadas para que se determine qual é a distribuição que o parâmetro possui com base nos conhecimentos do pesquisador ou especialista.

Três princípios devem ser considerados nesse método:

1. Fazer as perguntas para que sejam entendidas pelo pesquisador numa linguagem simples e familiar (termos técnicos).
2. Identificar os principais efeitos que influenciam o problema estudado.
3. Combinar as informações recebidas pelo questionário.

b) Método do histograma

Esse método é um processo iterativo no qual o pesquisador monta intervalos dos possíveis valores que o parâmetro possa ter e avalia (para cada intervalo) a probabilidade de esses valores ocorrerem. Com base nessa informação, o pesquisador monta um histograma (gráfico de barras) para estimar a densidade da probabilidade.

Processo iterativo: é a repetição do processo de estimação do parâmetro até o valor estimado convergir para um valor próximo ao valor real.

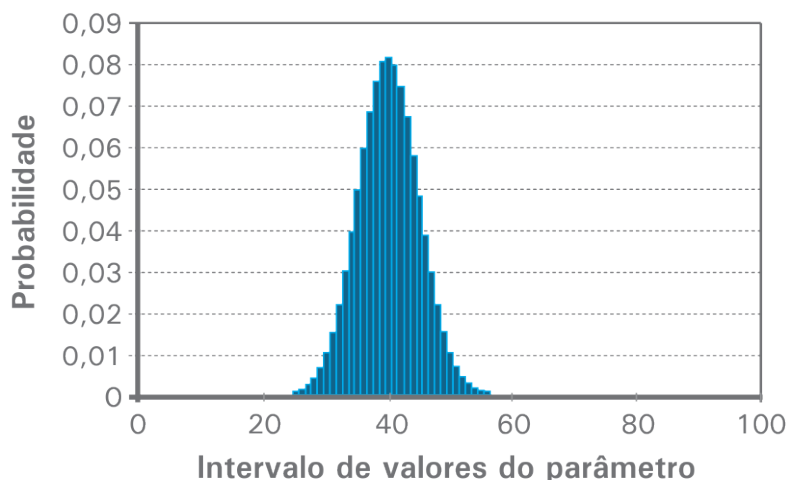


Figura 3. Histograma com os intervalos para os possíveis valores do θ estudado.
Fonte: Blasco (2008).

c) Método da verossimilhança

Esse método utiliza as verossimilhanças de possíveis valores dos parâmetros estudados e constrói, por meio de uma análise de regressão, uma função que passa por todos esses pontos. Essa função é, então, usada como *priori*.

Tipos de distribuição *a priori*

Priori própria

Nela há informação substancial sobre o parâmetro estudado, pois refere à informação contida em dados anteriores que são relevantes para o problema presente e que foi obtida de maneira científica ou resultante de considerações teóricas ou de opinião pessoal. Por exemplo, estimativas de herdabilidade de outros trabalhos podem ser, nesse caso, utilizadas na avaliação genética dos animais do seu trabalho (ASSIS, 2002; PAULINO et al., 2003).

Para que uma *priori* seja considerada como própria, a equação abaixo deve ser sempre igual a 1.

$$\int_a^b f(x)dx = 1$$

Dessa forma, a *distribuição a posteriori* (o resultado) vai ser uma distribuição de densidade de probabilidade na qual o valor do parâmetro pode ser estimado e a inferência realizada.

Quando a equação da densidade de probabilidade não for igual a 1, a *distribuição a posteriori* não será uma distribuição de densidade de probabilidade e não poderá realizar as inferências. A esse tipo de *priori* damos o nome de ***priori imprópria***.

A *priori* imprópria não é uma distribuição mas sim apenas um valor e assim não deve ser usada em uma análise bayesiana.

Priori não informativa ou vaga

Esse tipo de *priori* não dá nenhuma informação adicional (em relação ao conjunto de dados utilizado) sobre a distribuição do parâmetro estudado. Isto ocorre quando não se consegue extrair o conhecimento que essa *priori* oferece ou essa informação é pouco expressiva em relação às informações contidas nos dados (PAULINO et al., 2003). Assim, esse tipo de *priori* traz um conhecimento vago ou não informativo e a utilização dessa *priori* pode se tornar um problema bastante complexo na análise.

RESUMINDO

Priori não informativa = informação *a priori* insuficiente.

A primeira ideia de "não informação" está ligada ao fato de que todos os valores possíveis do parâmetro possuem a mesma probabilidade de ocorrer e esta *priori* será representada (Figura 4) por uma distribuição uniforme (BLASCO, 2008; EHLERS, 2003; SORENSEN; GIANOLA, 2002).

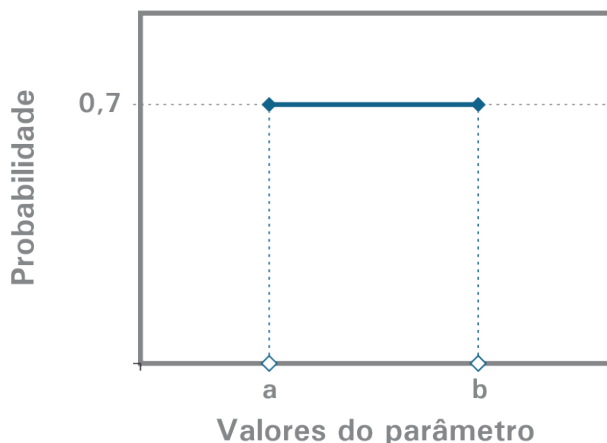


Figura 4. Esquema gráfico da curva de densidade de probabilidade de uma *priori* não informativa ou vaga.

Fonte: Blasco (2008).

No entanto, esse tipo de *priori* pode ser utilizada como ponto de referência na análise. A *priori* de referência é definida como uma função matemática que descreve a situação em que os dados têm maior influência no resultado e na inferência do que a informação *a priori*. O método mais utilizado, nesse caso, é a regra de Jeffreys (PAULINO et al., 2003).

Distribuições *a priori* usadas no melhoramento animal

Considere um modelo linear misto:

$$y = Xb + Zu + e$$

Onde:

y - é o vetor de observações.

X e Z_i - são as matrizes de incidências dos efeitos “fixos” e aleatórios, respectivamente.

b - é o vetor dos efeitos “fixos” desconhecidos.

u_i - é o vetor efeitos aleatórios.

e - é o vetor aleatório residual.

As probabilidades *a priori* de todos os efeitos do modelo e dos componentes de variância devem ser estabelecidas pelo melhorista. Usualmente, para a variável fixa X e o vetor b , assume-se uma distribuição uniforme, que indica ausência de conhecimento sobre esse vetor (CARNEIRO JUNIOR, 2001; POLICARPO, 2001).

Para os efeitos aleatórios, normalmente assume-se uma distribuição normal, mas para os componentes de variância σ_u^2 e σ_e^2 , que são variáveis aleatórias também, usualmente, a distribuição utilizada é a Qui-quadrado invertida (CARNEIRO JUNIOR, 2001). Para o vetor de valores genéticos aditivos, a , de acordo com a teoria da genética quantitativa, assume-se a distribuição normal com média zero e variância $A\sigma^2a$, em que A é a matriz de parentesco (POLICARPO, 2001).

No modelo animal, a distribuição da função de verossimilhança é uma distribuição normal com média $Xb + Za$ e variância $I\sigma^2e$. A distribuição *a priori* de θ tem que levar em conta a distribuição *a priori* dos seus diferentes componentes (b , a , σ^2a , σ^2e) (POLICARPO, 2001).

Inferência bayesiana

Variável aleatória

Define-se variável aleatória como um número real x_i associado ao evento E_i e esse número está ligado às características que descrevem o parâmetro estudado (YORIYAZ, 2010).

- a. **Variável aleatória discreta:** contém um número definido e exato de pontos amostrais. São observações contadas. Ex: sexo do animal (macho ou fêmea), diagnóstico positivo de prenhes em bovinos (sim ou não).
- b. **Variável aleatória contínua:** os pontos amostrais são contínuos, podendo assumir qualquer valor. São observações medidas. Ex: produção de leite, peso ao desmame, peso da carcaça.

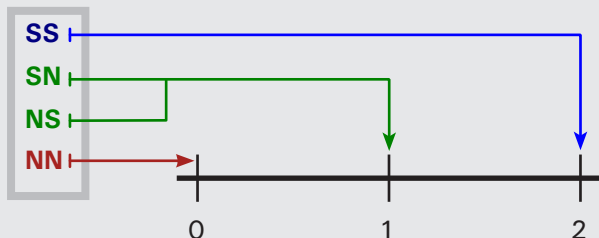
À medida que as iterações ocorrem, a variável aleatória assume valores definidos que vão ser futuramente utilizados na inferência por meio da descrição de suas probabilidades. Esses valores definidos devem estar dentro de um determinado limite (entre x_1 e x_2) e o valor real do parâmetro estudado deve estar nesse intervalo (YORIYAZ, 2010).

Exemplo:

EXPERIMENTO

- Uma vaca ter 2 diagnósticos positivos de prenhez consecutivos
- S = sim
- N = não

Evento



Assim, teremos para cada evento as seguintes probabilidades:

$$P(X = 0) = P(NN) = 25\%$$

$$P(X = 1) = P(SN \cup NS) = 50\%$$

$$P(X = 2) = P(SS) = 25\%$$

Densidade de Probabilidade

Na estatística, a distribuição de densidade de probabilidade é uma função utilizada para representar a distribuição de probabilidade de uma variável aleatória contínua. Utiliza o cálculo de integral como ferramenta.

$$f(x) = \int_a^b f(x)dx$$

Ao fazer inferências sobre variáveis contínuas, existe um número bastante grande de possíveis valores que o parâmetro estudado pode ter. No entanto, cada valor tem uma probabilidade de ocorrer e assim (para facilitar a inferência) utiliza-se uma função, chamada de **Função Densidade de Probabilidade**, como auxílio na avaliação dos resultados.

Essa função, depois de calculada a sua integral, mostra qual valor possui a maior probabilidade de ocorrer (em forma de gráfico) dentre todos os valores possíveis que o parâmetro pode apresentar. Em outras palavras, é a probabilidade de um determinado valor do parâmetro estar entre dois valores A e B (BLASCO, 2008).

OBSERVAÇÃO

Somente serão utilizadas na inferência as probabilidades relacionadas aos valores do parâmetro que estão dentro do intervalo.

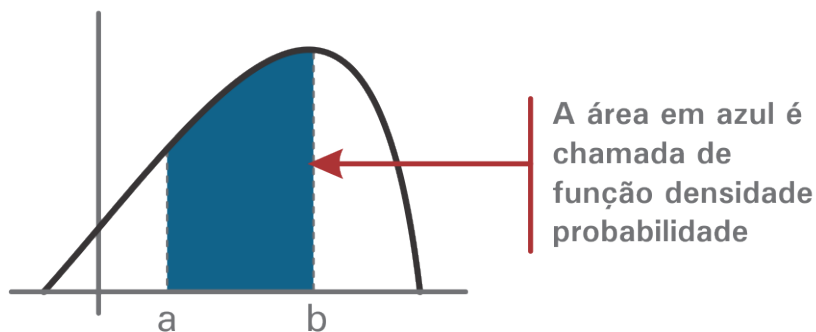


Figura 5. Gráfico representando uma função de densidade de probabilidade.
Fonte: Blasco (2008).

Intervalo de credibilidade

O intervalo de credibilidade entre os pontos a e b (chamado pelos estatísticos clássicos de intervalo de confiança) contém o verdadeiro valor do parâmetro, com uma probabilidade de 95% e que são facilmente construídos pelos métodos iterativos usados na estatística bayesiana (BLASCO, 2008).

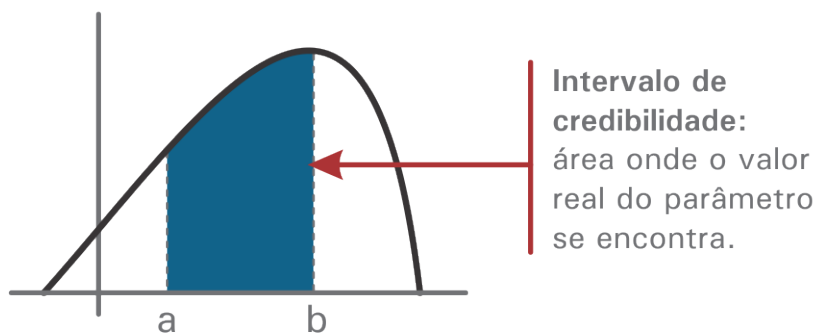


Figura 6. Gráfico representando o intervalo de credibilidade.
Fonte: Blasco (2008).

Distribuição *a posteriori*

Do ponto de vista bayesiano, fazer a inferência da análise é utilizar a distribuição *a posteriori*. Na abordagem bayesiana, a forma de estudar os parâmetros consiste em esboçar o gráfico da distribuição *a posteriori* que mostra o intervalo no qual o valor real do parâmetro está e calcula as características dessa distribuição (média, variância, moda, etc.), obtendo assim toda a informação desse parâmetro (LEANDRO, 2001).

Inferência é a passagem da hipótese para a conclusão.

A distribuição *a posteriori* é a descrição completa e a atualização do conhecimento sobre o parâmetro obtido a partir da informação *a priori* e da informação dos dados (PAULINO et al., 2003). Ou seja, o pesquisador observa os dados após o experimento e esses dados são usados para modificar as probabilidades anteriores dos valores do parâmetro, produzindo um conjunto de probabilidades posteriores ao experimento (SALSBURG, 2009).

PROBABILIDADE ANTERIOR → DADOS → PROBABILIDADE POSTERIOR

Na prática, faz-se o resumo da informação sobre o parâmetro (que agora está atualizada) de modo que os aspectos interessantes (em termos de probabilidade) desse parâmetro possam ser observados (PAULINO et al., 2003).

A distribuição *a posteriori* apresenta a localização, dispersão e distribuição dos possíveis valores do parâmetro.

Na estimação da distribuição *a posteriori*, pode-se adotar dois critérios para realizar as inferências, que também podem ser usados conjuntamente: a probabilidade da *posteriori* propriamente dita e o Fator de Bayes (LEANDRO, 2001; PAULINO et al., 2003).

Distribuições que compõem a distribuição *a posteriori*

Distribuição de probabilidade conjunta

Há situações em que precisamos avaliar dois parâmetros simultaneamente, no caso do melhoramento genético, por exemplo, e as probabilidades associadas aos valores possíveis para cada parâmetro são desconhecidas (separadamente). Portanto, é possível representar a probabilidade para a ocorrência dos dois eventos pela distribuição de probabilidade conjunta $f(x,y)$ e realizar as inferências em conjunto para as duas características estudadas.

Distribuição de probabilidade marginal

Na estatística bayesiana, as inferências sobre os parâmetros avaliados da característica (por exemplo, produção de leite) são feitas por meio da distribuição *a posteriori*, que contém toda a informação dos parâmetros envolvidos na expressão dessa característica (por exemplo, a herdabilidade pode ser dividida em vários componentes: variância aditiva, variância do efeito materno, etc.).

Porém, às vezes, faz-se necessário estudar apenas um desses parâmetros x (no exemplo anterior, variância aditiva). Então, para obter as informações desse efeito particular, deve-se integrar a distribuição *a posteriori* obtida como resultado em relação a todos os outros efeitos que afetam a característica avaliada.

Assim, a partir da distribuição conjunta de todos os efeitos ou parâmetros estudados que compõem a característica, podemos obter a distribuição de densidade de probabilidade marginal de x e interpretar o efeito de interesse isoladamente dos outros efeitos, como por exemplo, estudar o efeito da variância aditiva incluída no valor genético obtido para produção de leite (BLASCO, 2011; PETERNELLI, 2011).

Distribuição de probabilidade condicional

A distribuição condicional dos dados é assim chamada porque a probabilidade de um dado valor do parâmetro ocorrer está condicionada aos valores da média e da variância obtidos pelos dados estudados.

Condicionada: dependente de uma condição imposta.

Assim, para cada valor diferente de média e variância, a distribuição observada, como resultado, muda e consequentemente a inferência realizada também (BLASCO, 2008).

Obtenção do *posteriori* na estatística bayesiana

Método de Monte Carlo via cadeia de Markov (MCMC)

O propósito deste item é apresentar alguns elementos sobre a teoria do método Monte Carlo via cadeia de Markov (MCMC) para que o leitor possa entender os mecanismos utilizados na obtenção das distribuições *a posteriori* pela estatística bayesiana.

A distribuição *a posteriori* do parâmetro estudado contém todas as informações necessárias para se fazer as inferências a respeito desse parâmetro. A partir dessas informações, é possível integrar a distribuição *a posteriori* $[p(\theta/\mathbf{x})]$ para esboçar o gráfico dessa distribuição por meio dos valores numéricos que resumem a informação contida nos dados. Infelizmente, esse procedimento, na maioria dos casos, é muito difícil de ser feito, inviabilizando a aplicação da estatística bayesiana (POLICARPO, 2001).

Então, como resolver este problema?

A solução foi utilizar métodos de inferência baseados na simulação iterativa. No caso da inferência bayesiana, o **Método de Monte Carlo via cadeia de Markov (MCMC)** (LEANDRO, 2001; PAULINO et al., 2003).

Simulação refere-se a qualquer método cuja intenção é imitar algum método de análise estatístico sem a necessidade de cálculos complexos, como por exemplo, a integração.

Esse método já era conhecido há séculos pelos estatísticos, mas começou a ser utilizado efetivamente somente nas últimas décadas do século 20. O nome Monte Carlo (capital de Mônaco conhecida como a cidade mundial dos jogos de azar) foi inspirado no interesse do pesquisador Nicholas Metropolis, que desenvolveu os algoritmos para implementação desse método estatístico baseado nas observações dos jogos de azar de Stanislaw Ulam (YORIYAZ, 2010).

Atualmente o método de Monte Carlo é utilizado como ferramenta no auxílio à pesquisa nos mais variados campos, por exemplo, o desenho de reator nuclear, como dinâmica quântica, terapia de câncer por radiação, fluxo de tráfego, evolução estelar, agropecuária e econometria.

A simulação de Monte Carlo é realizada por meio da amostragem aleatória das densidades de probabilidades *a posteriori* conjuntas ou marginais e, então, a média, o desvio-padrão e os intervalos de confiança podem ser estimados diretamente dessas amostras, sem que seja necessária a resolução de integrais (ASSIS, 2002).

Os números podem ser obtidos por meio de algum processo aleatório (tabelas, roletas, etc.) ou diretamente do computador, por meio de funções ou algoritmos específicos, como os utilizados na bayesiana (Amostrador de Gibbs e Metropolis-Hasting).

Os modelos de simulação (ou amostragem) são utilizados para analisar uma decisão envolvendo risco, ou seja, um modelo que verifica o comportamento de um ou mais efeitos estudados que não são conhecidos com certeza. Nesse caso, esses efeitos são conhecidos como variável aleatória e o seu comportamento são descritos por uma distribuição de probabilidade.

Nesse método, o valor estimado é uma média dos valores obtidos em todas as iterações realizadas e não se pode afirmar que este seja o valor exato do parâmetro.

Iteração: tornar a fazer, repetição.

Essa diferença entre os valores estimado e real é chamada de “erro de Monte Carlo”. A confiabilidade da estimativa do parâmetro estudado será maior quanto menor for o erro e será tanto melhor quanto maior for o número de amostras utilizadas na estimação (ROSA, 1998; SAVIAN, 2008; TIMPANI, 2011).

O método da Cadeia de Markov é um processo em que a probabilidade de certo valor no tempo futuro depende do valor do parâmetro na iteração atual, mas não dos valores estimados em iterações anteriores. Esses valores gerados pela cadeia de Markov são considerados amostras aleatórias da distribuição de probabilidade do parâmetro estudado (SILVA, 2006; TIMPANI, 2011).

A ação conjunta dos dois métodos é chamada de método Monte Carlo via Cadeia de Markov (MCMC), usado, frequentemente, na estimação de parâmetros genéticos, por meio da estatística bayesiana no melhoramento genético animal (SILVA, 2006; TIMPANI, 2011).

O Método MCMC é, portanto, um modelo de simulação que utiliza a geração de uma sequência (Cadeia de Markov) de números aleatórios (amostras) para atribuir valores aos parâmetros que se deseja investigar.

A ideia dos processos iterativos, como o Monte Carlo, é gerar valores de variável aleatória x , em que, a cada passagem (iteração), milhares de valores vão sendo gerados para o parâmetro, criando uma distribuição para esse parâmetro, até chegar a uma distribuição bastante próxima da original, ou seja, quando ocorre uma convergência. Ao gerar uma sequência com muitos valores simulados, os valores poderão ser

usados para interpretar as características da distribuição *a posteriori* obtida no experimento (LEANDRO, 2001; POLICARPO, 2001).

O processo pode ser pensado como um sistema que se move aleatoriamente por meio de “estados”, sem ter lembrança de onde ele esteve anteriormente. Em outras palavras, a cadeia gerada vai gradualmente “esquecendo” o valor inicial que foi dado pelo pesquisador (LEANDRO, 2001).

Para que ocorra esse “movimento da cadeia” são necessários dois “instrumentos de trabalho”: a probabilidade de transição e a distribuição de probabilidade inicial.

Probabilidade de transição é a probabilidade de um valor gerado pela cadeia ser aceito ou rejeitado pelo sistema. Assim, a transição do estado x para o estado y ocorre de acordo com a probabilidade $p(x, y)$.

A distribuição de probabilidade inicial pode ser resumida no valor inicial do parâmetro estudado.

O valor gerado na iteração só depende do valor da iteração anterior (LEANDRO, 2001). Assim, o valor gerado para o parâmetro em uma iteração só é relevante para gerar o valor da iteração seguinte, sendo, portanto, irrelevante na predição dos valores futuros (ASSIS, 2002).

Isto gera um problema de ordem prática: os valores consecutivos são altamente correlacionados, interferindo nas estimativas dos parâmetros estudados (CARNEIRO JUNIOR, 2001; TIMPANI, 2011). Dessa forma, se o primeiro valor for utilizado como resultado, é necessário que ocorra o intervalo amostral (*thining interval*), no qual alguns valores após o valor utilizado são descartados para não prejudicar os resultados.

Essa simulação iterativa é menos eficiente que o cálculo direto, mas é aplicável em praticamente todos os casos e desempenha um papel extremamente marcante na inferência bayesiana. Sua principal falha

está ligada à convergência, para a qual até o presente momento não foram ainda desenvolvidos métodos eficientes para determiná-la (LEANDRO, 2001).

Os principais componentes na simulação por MCMC são as densidade de probabilidade e os números aleatórios gerados na simulação, enquanto o processo de simulação é feito por meio de técnicas de amostragem.

Uma grande vantagem dos métodos de simulação é que, após uma amostra de valores aleatórios ser gerada, pode-se facilmente calcular características das distribuições marginais e condicionais e, a partir delas, a distribuição *a posteriori* dos parâmetros estudados (como por exemplo, os valores das variâncias genética, ambiental e residual para a característica ganho de peso em codornas de corte).

Algoritmo Metropolis-Hastings

Nesse algoritmo, a cada iteração realizada no processo, é escolhido um valor inicial (um ponto candidato y) para os parâmetros da distribuição *a posteriori* (LEANDRO, 2001).

Se esse ponto candidato for aceito pelo sistema como sendo o valor aproximado do parâmetro estudado, o processo de simulação continua ocorrendo até a convergência. Caso contrário, o processo para e um novo valor inicial deve ser utilizado (LEANDRO, 2001).

Para que o sistema identifique e aceite o valor gerado, no algoritmo Metropolis-Hastings, obtém-se para cada valor dado a probabilidade de transição para o próximo estado (CARNEIRO JUNIOR, 2001). Um sistema exerce estado de transição (e isso ocorre com certa probabilidade), quando ocorre a troca do valor do parâmetro de uma iteração (ou estado) para outro valor, na iteração seguinte.

Amostrador de Gibbs

O Amostrador de Gibbs é um algoritmo do tipo Metropolis-Hastings, da família do método Monte Carlo da Cadeia de Markov (POLICARPO, 2001) e o mais utilizado no melhoramento genético animal.

O Amostrador de Gibbs é um procedimento que obtém amostras da distribuição conjunta ou marginal de todos os parâmetros estudados (por exemplo, os componentes de variância que descrevem uma característica) por meio de repetidas amostragens, formando uma Cadeia de Markov. Essa cadeia corresponde a uma sequência de amostras de possíveis valores atribuídos aos parâmetros estudados (POLICARPO, 2001).

À medida que o número de iterações aumenta, a sequência de valores gerados (cadeia) aproxima-se da distribuição de equilíbrio, ou seja, da densidade marginal e, quando esse equilíbrio é atingido, diz-se que ocorreu a convergência (SAVIAN, 2008; SILVA, 2006; TIMPANI, 2011).

Para iniciar o processo de amostragem, um valor inicial de Y_0 é assumido e os demais valores da cadeia são gerados de forma iterativa a partir das distribuições condicionais dos parâmetros (POLICARPO, 2001).

Assim, a ideia desse algoritmo é gerar uma cadeia única e longa a partir de um estado inicial $[\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})]$ e com comprimento $m = l + kn$, em que m é o número de iterações; l é o número inicial de iterações necessárias para que a cadeia atinja a convergência; n é o número de iterações que vão ser utilizadas na aplicação do MCMC e k é o intervalo amostral utilizado para eliminar a correlação entre as iterações sucessivas (SAVIAN, 2008; TIMPANI, 2011).

No processo de amostragem de Gibbs, o número de total de iterações que a análise deve realizar é determinado pelo número de iterações necessárias para a convergência, para o descarte inicial (*burn-in*) e o intervalo amostral (*thinning interval*) (TIMPANI, 2011).

A convergência da cadeia para a distribuição de equilíbrio não ocorre imediatamente após o início do processo iterativo e normalmente as primeiras amostras (valores) geradas pelo Amostrador são muito correlacionadas com os valores dados inicialmente pela distribuição *a priori* utilizada. Então, as amostras iniciais devem ser descartadas na estimação final dos parâmetros. O tamanho desse período de descarte inicial ou *burn-in* é frequentemente determinado pela inspeção visual de gráfico dos valores gerados ao longo de cada iteração (POLICARPO, 2001).

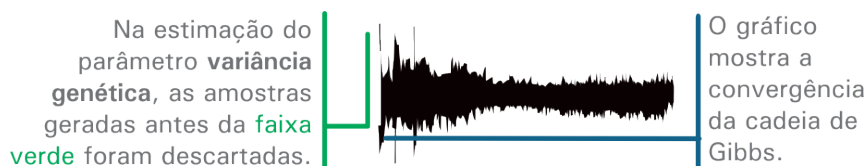


Figura 7. Diagnóstico de convergência para variância genética da característica ganho de peso em codornas europeias.

Fonte: Timpani (2011).

Seja $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$ um valor arbitrário inicial para o vetor θ .

Procede-se iterativamente da seguinte forma:

1) obtém-se $\theta_1^{(1)}$ de $p(\theta_1 | x, \theta_2^{(0)}, \dots, \theta_k^{(0)})$,

obtém-se $\theta_2^{(1)}$ de $p(\theta_2 | x, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$,

obtém-se $\theta_3^{(1)}$ de $p(\theta_3 | x, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$,

...

obtém-se $\theta_k^{(1)}$ de $p(\theta_k | x, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_{k-1}^{(1)})$.

Completa-se, assim uma iteração do esquema e uma transição de

θ^0 para $\theta^1 = (\theta_1^{(1)}, \dots, \theta_k^{(1)})$.

2) O esquema anterior é repetido com θ^1 , obtido anteriormente, como vetor inicial para obter um novo vetor θ^2 e, assim, haver uma transição de θ^1 para $\theta^2 = (\theta_1^{(2)}, \dots, \theta_k^{(2)})$.

3) Repete-se m vezes este ciclo de geração de observações aleatórias de cada uma das distribuições condicionais, produzindo, assim, $\theta^0, \dots, \theta^m$.

Figura 8. Esquema ilustrativo do Algoritmo do Amostrador de Gibbs.

Fonte: Blasco (2008).

Diagnóstico de convergência

Os métodos de MCMC são uma ótima ferramenta para resolução de muitos problemas práticos na metodologia bayesiana. Porém, as questões ligadas à convergência da cadeia ainda são uma incógnita na utilização dessa metodologia (ABANTO-VALLE, 2005).

A convergência da cadeia de Markov ocorre quando se percebe certa estabilidade, ou seja, os valores gerados no processo se mantêm os mesmo de iteração para iteração (LEANDRO, 2001; POLICARPO, 2001).

Para verificar a convergência, é necessário examinar certas características da cadeia produzida. Nesse caso, existem vários métodos (estatísticos e gráficos) que podem ser utilizados com esse fim. No entanto, dois pontos devem ser considerados na escolha do método de avaliação da convergência (LEANDRO, 2001):

1. Nenhum método é infalível e a recomendação da convergência pode ser feita mesmo quando a cadeia ainda não se estabilizou.
2. É conveniente para uma maior acurácia do processo que vários métodos sejam usados em conjunto na determinação da convergência.

Uma importante consideração a respeito da convergência é que existe uma única distribuição de equilíbrio (lembrem-se: quando a cadeia atinge a convergência é porque ela se estabilizou) e duas propriedades são responsáveis por isso: aperiocidade e irredutível (SORENSEN; GIANOLA, 2002).

Aperiocidade: essa propriedade garante que a cadeia não fique analisando um número finito e determinado de valores do parâmetro (ABANTO-VALLE, 2005).

Irredutível: para uma cadeia ter essa propriedade é necessário que ela possa “visitar” todos os possíveis conjuntos de valores do parâmetro, independente do valor inicial dado (ABANTO-VALLE, 2005).

Para avaliar a convergência dos métodos de MCMC, faz-se uso de alguns métodos já empregados na literatura. As técnicas mais populares são Geweke (1992) e Heidelberger e Welch (1983), que usam resultados baseados em análise espectral; Raftery e Lewis (1992), que permite calcular quantas iterações são necessárias para uma cadeia atingir a distribuição estacionária por meio da estimação de *quantis* posteriores com uma precisão previamente fixada; Gelman e Rubin (1992), que usa resultados baseados na análise de variância clássica para duas ou mais cadeias simuladas com valores iniciais diferentes. Esses métodos e outros foram comparados no trabalho de Cowles e Carlin (1996), que concluíram que não se pode afirmar qual deles é o mais eficiente (ABANTO-VALLE, 2005).

Crítérios de seleção do modelo e hipóteses para análise estatística

Até o momento, observou-se que a distribuição *a priori* é deduzida e a função de verossimilhança está de acordo com o que foi apresentado pelos dados, a distribuição *a posteriori* é obtida e as inferências sobre os parâmetros estudados (como, por exemplo, efeito genético, efeito materno, desvio de dominância, genes de efeito principal, etc.) podem ser realizadas.

No entanto, o modelo ou as hipóteses adotadas para análise devem ser “boas o suficiente” para evitar que a incerteza sobre as informações seja muito grande e comprometa a veracidade dos resultados.

Sempre há alternativas que podem ou não ser adotadas.

No melhoramento animal, é de fundamental importância na análise estatística (bayesiana ou clássica) a determinação dos efeitos (como, por exemplo, genético, materno, dominância, genes de efeito principal, etc.) que influenciam as características em estudo.

O uso do modelo errado ou incompleto pode trazer consequências para a acurácia das estimativas dos valores genéticos, parâmetros genéticos ou componentes de variância. Assim, a escolha do modelo apropriado para a análise é uma atividade fundamental, que vem tornando-se cada vez mais importante na análise estatística, uma vez que, em razão dos avanços computacionais, é possível construir modelos cada vez mais complexos.

Uma das técnicas mais usadas com esse fim, quando a estatística bayesiana é aplicada no melhoramento animal, é o Fator de Bayes.

Fator de Bayes

O Fator de Bayes (FB) é usado para comparar a probabilidade *a posteriori* ao testar duas hipóteses diferentes. Em outras palavras, corresponde à relação entre as probabilidades das diferentes *prioris* que podem ser utilizadas no experimento.

O Fator de Bayes é assim definido, segundo Kass e Raftery (1995):

$$FB_{ij} = \frac{P(y|M_i)}{P(y|M_j)}$$

Em que:

$P(y|M_i)$ é a probabilidade da *posteriori* que expressa a premissa representada pela hipótese alternativa (H_1).

$P(y|M_j)$ é a probabilidade da *posteriori* que expressa a premissa representada pela hipótese alternativa (H_0).

Como exemplo, considerando que você queira determinar a existência de genes de efeito principal (GEP) na característica Idade ao primeiro parto em bovinos da raça Nelore (TIMPANI, 2011).

O GEP é um gene que tem uma grande influência sobre a expressão da característica quantitativa.

As hipóteses a serem traçadas para o problema são:

- H_0

= a característica IPP não possui GEP com probabilidade *a priori* p .
- H_1

= a característica IPP possui GEP com probabilidade *a priori* q .

$$FB_{ij} = \frac{H_1}{H_0}$$

A interpretação do FB é bastante simples e direta. Se $FB > 1$, tem-se a indicação de que a hipótese que está no numerador (corresponde a H_1) é a verdadeira; caso contrário, se $FB < 1$, a hipótese que está no denominador (corresponde a H_0) é a verdadeira (SILVA, 2006; TIMPANI, 2011).

Existem outras interpretações para o Fator de Bayes que podem ser utilizadas para seleção de modelos. Na Tabela 2, é apresentada a interpretação segundo Kass e Raftery (1995).

Tabela 2. Fator de Bayes, segundo Kass e Raftery (1995).

Fator de Bayes	Evidência contra hipótese nula (H_0)
0,0 a 2,0	Baixa
2,1 a 5,0	Positiva
5,1 a 10,0	Forte
> 10,0	Decisiva

Essa interpretação baseia-se na evidência proveniente dos dados contra uma das hipóteses, que segundo Kass e Raftery (1995) é a hipótese nula (H_0).

A evidência baixa indica que a hipótese nula (H_0) é a mais provável de ser a verdadeira. A evidência positiva mostra que a hipótese alternativa (H_1) é a mais provável de ser a correta. As evidências forte e decisiva confirmam que a hipótese alternativa (H_1) é a correta.

Outro exemplo: avaliando a tabela 3, é possível concluir pelas probabilidades *a posteriori* e pelo Fator de Bayes obtido que a característica ganho de peso aos 345 dias (GP345) na raça Nelore apresentou evidência de possuir GEPs (TIMPANI, 2011).

Tabela 3. Probabilidade *a posteriori* e Fator de Bayes para o número de genes de efeito principal (GEP) na característica ganho de peso aos 345 dias.

Probabilidade <i>a posteriori</i>						Fator de Bayes				
0	1	2	3	4	5	1/0	2/1	3/2	4/3	4/5
0	0	0,98	0,02	0	0	na	29,0	-5,3	na	na

Pela análise da probabilidade *a posteriori* apresentada na Tabela 3, observa-se na característica GP345, que a maior probabilidade está relacionada à presença de dois genes de efeito principal influenciando sua expressão fenotípica.

ATENÇÃO

Isto é uma evidência, uma probabilidade de que a hipótese seja verdadeira na prática.

A interpretação do FB (Tabela 3) é baseada na confrontação de duas hipóteses (H_0 e H_1) relacionadas às probabilidades de os diferentes números de GEPs estarem influenciando essa característica.

Analisando o Fator de Bayes (Tabela 3), observa-se que há uma evidência decisiva (acima de 10,0) de que GP345 seja influenciada por dois GEPs.

Isto é verificado na comparação entre as hipóteses nula ($H_0: N_{\text{GEP}'} = 1$) e alternativa ($H_1: N_{\text{GEP}'} = 2$), que possui maior Fator de Bayes, e na comparação entre as hipóteses nula ($H_0: N_{\text{GEP}'} = 2$) e alternativa ($H_1: N_{\text{GEP}'} = 3$), em que o valor obtido para o Fator de Bayes é negativo, indicando que o denominador tem o número correto de GEPs para GP345.

Portanto, a conclusão a que chegamos é que pode haver dois GEPs influenciando a expressão fenotípica da característica GP345 em bovinos da raça Nelore.

Princípios da estatística bayesiana no modelo animal

Modelo animal

Na estatística bayesiana, não existe distinção entre variáveis fixas e aleatórias e, portanto, os efeitos considerados fixos no melhoramento animal (por exemplo, grupo de contemporâneos) são usados como efeitos aleatórios sobre os quais se tem pouco ou nenhum conhecimento *a priori* (GIANOLA; FERNANDO, 1986; POLICARPO, 2001).

Outro ponto interessante na aplicação da estatística bayesiana no melhoramento animal, segundo Gianola e Fernando (1986), é que a distribuição dos dados não precisa ser necessariamente uma distribuição Normal, sendo possível utilizá-la na avaliação de características com outros tipos de distribuição (como por exemplo, características de limiar: diagnóstico positivo de prenhes de vacas).

Quando Henderson propôs o BLUP como método de avaliação genética (que é atualmente o método padrão utilizado pelo melhoramento animal) considerou situações e populações em que as variâncias e covariâncias genética e residual eram conhecidas e observou que pouco sucesso é

alcançado nos métodos de seleção quando as variâncias precisam ser estimadas na própria população sob seleção (POLICARPO, 2001).

No entanto, no caso da metodologia bayesiana, as variâncias não precisam ser conhecidas e podem ser estimadas a partir dos dados das populações selecionadas. Nesse caso, a distribuição *a posteriori* dos valores genéticos e parâmetros genéticos é a mesma com ou sem seleção ou acasalamento controlado. Assim, a estatística bayesiana pode fornecer inferência a respeito dos valores e parâmetros genéticos mesmo na presença de seleção (POLICARPO, 2001), o que pode ser um fator de viciamento das estimativas na clássica.

Sorensen et al. (1994), comparando a metodologia clássica com a bayesiana em relação à resposta à seleção, observaram que há diferenças entre os resultados obtidos nas duas metodologias, quando o volume de dados é pequeno. Porém, quando o conjunto de dados é grande, não houve diferença entre os resultados. Isto ocorre porque os dados são muito informativos em relação aos parâmetros analisados e a função de verossimilhança domina as informações *a priori*, e vice-versa. Uma das grandes vantagens dos métodos bayesianos é que eles permitem verificar conclusões falsas baseadas em verossimilhanças fracas ou pouco informativas.

Estimação dos componentes de variância

No melhoramento animal, as variâncias e covariâncias genéticas não são conhecidas e devem ser estimadas.

Assumindo que a distribuição *a priori* para os efeitos fixos e componentes de variância não são informativas, na metodologia clássica, os componentes de variância são obtidos como a moda da distribuição *a posteriori* conjunta após integração em relação aos efeitos fixos. Assim, essas estimativas são uma aproximação das distribuições marginais dos componentes de variância em relação aos efeitos fixos incluídos no modelo (POLICARPO, 2001). Essa distribuição

marginal é particularmente importante na inferência, uma vez que o objetivo da avaliação é estimar os componentes de variância de alguns dos parâmetros apenas e que pela estatística bayesiana podem ser diferenciados.

Estimação dos valores genéticos e seleção

Os dados de campo, cuja avaliação genética é realizada, usualmente provêm de rebanhos nos quais a seleção é rotineiramente praticada. A questão de como estimar valores genéticos em populações submetidas à seleção é de grande importância prática, pois, na metodologia clássica o fato de a população estar em seleção pode ser um problema para a obtenção de estimativas acuradas.

No contexto bayesiano, pode-se mostrar que as inferências sobre os valores genéticos e parâmetros genéticos podem ser feitas a partir da distribuição *a posteriori* obtida como resultado, como se a seleção não tivesse ocorrido na população avaliada, ou seja, concluindo, a distribuição *a posteriori* dos parâmetros genéticos e dos valores genéticos, pela metodologia bayesiana, é a mesma com ou sem seleção ou acasalamento não aleatório (ASSIS, 2002).

Referências

ABANTO-VALLE, C. A. A. **Métodos de simulação estocástica em modelos dinâmicos não lineares**: uma aplicação em modelos de volatilidade. 2005. 189 f. Tese (Doutorado em Estatística) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.

ASSIS, G. M. L. **Predição bayesiana e clássica aplicada ao melhoramento animal**. 2002. 40 f. Exame de qualificação (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.

BAYES, T. An essay towards solving a problem in the doctrine of chances. **Philosophical Transactions of The Royal Society**, v. 53, p. 370–418, 1763. (Reprinted in **Biometrika**, v. 45, n. 3/4, p. 296–315, 1958).

BLASCO, A. The bayesian controversy in animal breeding. **Journal of Animal Science**, Champaign, v. 79, n. 3, p. 2023-2046, Mar. 2001.

BLASCO, A. **An easy introduction to Bayesian statistics with MCMC**: lecture notes. Lavras, 2008.

CARNEIRO JUNIOR, J. M. **Uso de Métodos Bayesianos no Melhoramento Animal, com Ênfase na Amostragem de Gibbs**. 2001. 50 p. Exame de qualificação (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.

DICIONÁRIO Houaiss da língua portuguesa. [Rio de Janeiro]: Objetiva, 2009. Disponível em: <http://atlas.sct.embrapa.br/houaiss2009/cgi-bin/houaissnet.exe?func=historico&palavra=infer%EAncia:67738>. Acesso em: 29 abr. 2015.

EHLERS, R. S. **Introdução a inferência Bayesiana**. 2003. Disponível em: <http://www.leg.ufpr.br/~paulojus/CE227/ce227/ce227.html>. Acesso em: 18 jan. 2011.

FARIA, C. U. de; MAGNABOSCO, C. de U.; REYES, A. de los; LÔBO, R. B.; BEZERRA, L. A. F. Inferência bayesiana e sua aplicação na avaliação genética de bovinos da raça Nelore: revisão bibliográfica. **Ciência Animal Brasileira**, v. 8, n. 1, p. 75-86, 2007.

FERREIRA, D. F. **Estatística básica**. Lavras: Editora UFLA, 2005. 664 p.

GIANOLA, D.; FERNANDO, R. L. Bayesian Methods in Animal Breeding Theory. **Journal of Animal Science**, v. 63, n. 1, p. 217-244, 1986.

KASS, R. E.; RAFTERY, A. E. Bayes factors. **Journal of the American Statistical Association**, New York, v. 90, n. 3, p. 773-795, June 1995.

LEANDRO, R. A. Introdução à estatística bayesiana: notas. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 9., 2001, Piracicaba. **Anais**: resumos. Piracicaba: ESALQ, 2001.

MEMÓRIA, J. M. P. **Breve História da Estatística**. Brasília, DF: Embrapa Informação Tecnológica: Embrapa-Secretaria de Gestão e Estratégia, 2004. 111 p. (Embrapa-Secretaria de Gestão e Estratégia. Texto para discussão, 21).

NOGUEIRA, D. A.; SAFÁDI, T.; BEARZOTI, E.; BUENO FILHO, J. S. de S. Análises clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência e Agrotecnologia**, Lavras, Ed. Esp., p. 1614-1624, dez. 2003.

PAULINO, C. D.; TURKMAN, A. A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.

PERERNELLI, L. A. **Programa analítico**. [Viçosa: UFV], 2011. Cap. 4, p. 47-67. Disponível em: <http://www.dpi.ufv.br/~peternelli/inf162.www.16032004/materiais/CAPITULO4.pdf>. Acesso em: 31 jan. 2011.

POLICARPO, A. **Estatística bayesiana aplicada ao melhoramento genético animal (Amostrador de Gibbs e suas aplicações)**. 2001. 17 f. Exame de qualificação (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.

ROSA, G. J. M. **Análise bayesiana para modelos lineares mistos robustos via Amostrador de Gibbs**. 1998. 57 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

SALSBURG, D. **Uma senhora toma chá...como a estatística revolucionou a ciência do século XX**. Rio de Janeiro: Editora Zahar, 2009. 286 p.

SAVIAN, T. V. **Análise bayesiana para modelos de degradabilidade ruminal**. 2008. 81 f. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

SILVA, F. F. **Análise bayesiana do modelo auto-regressivo para dados em painel: aplicação na avaliação genética de touros da raça Nelore**. 2006. 100 f. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

SORENSEN, D. A.; WANG, C. S.; JENSEN, J.; GEANOLA, D. Bayesian Analysis of Genetic Change Due to Selection Using Gibbs Sampling. **Genetics, Selection, Evolution**, v. 26, n. 4, p. 333-360, 1994.

SORENSEN, D.; GIANOLA, D. **Statistics for biology and health**. New York: Springer, 2002. 740 p.

TIMPANI, V. D. **Arquitetura genética de características de crescimento e reprodução em bovinos da raça Nelore**. 2011. 104 f. Tese (Doutorado em Zootecnia) - Universidade Federal de Lavras, Lavras.

YORIYAZ, H. **Fundamentos do Método de Monte Carlo para Transporte de Radiação**. 2010. 91 p. Apostila da disciplina.



Amazônia Oriental

Ministério da
**Agricultura, Pecuária
e Abastecimento**

GOVERNO FEDERAL
BRASIL
PÁTRIA EDUCADORA

CGPE 12080